R. K. Rai, M. Balmer, M. Rieser, V. S. Vaze, S. Schönfelder,
K. W. Axhausen

Capturing human activity spaces: New geometries

Rai, R. K., Balmer, M., Rieser, M., Vaze, V. S., Schönfelder, S., & Axhausen, K. W. (2007). Capturing Human Activity Spaces. Transportation Research Record: Journal of the Transportation Research Board, 2021(1), 70–80. https://doi.org/10.3141/2021-09

Capturing human activity spaces: New geometries

2006-11-15

R.K. Rai
IIT Guwahati
Guwahati 781039, Assam, India
Phone: +91-986-420-23-53
Email: rkrai@iitg.ernet.in

Michael Balmer
Institute for Transport Planning and Systems (IVT), ETH Zurich
8093 Zurich, Switzerland
Phone: +41-44-633-27-80
Fax: +41-44-633-10-57
Email: balmer@ivt.baug.ethz.ch

Marcel Rieser
VSP, TU-Berlin
10587 Berlin, Germany
Phone: +49-30-314-25-258
Fax: +49-30-314-26-269
Email: rieser@vsp.tu-berlin.de

V.S. Vaze
IIT Bombay
Powai, 400057 Mumbai, India
Email: vikrantv@mit.edu

Stefan Schönfelder
TRAFICO Verkehrsplanung
Fillgradergasse 6/2, 1060 Vienna, Austria
Phone: +43-1-586-4181
Fax: +43-1-586-41-8110
Email: stefan.schoenfelder@trafico.at

Kay W. Axhausen
Institute for Transport Planning and Systems (IVT), ETH Zurich
8093 Zurich, Switzerland
Phone: +41-44-633-39-43
Fax: +41-44-633-10-57
Email: axhausen@ivt.baug.ethz.ch

Words:	6234
Figures/Tables:	6/1
Total:	7984

ABSTRACT

Activity space, defined as “the local areas within which people move or travel during the course of their activities during a specified time period”, is a measure of an individual’s spatial behavior which captures individual and environmental differences and offers an alternative approach to studying the spatial reach of travelers. The shape and area of activity space is a product of how it is conceptualized and measured. This paper enlarges the set of geometries which can be used to describe activity space. It tests four parametric geometries (ellipse, superellipse, Cassini oval, and bean curve), which are identified as those capturing a specific share of all locations visited, i.e. 95%, while minimizing the area covered. They are estimated for a number of long-duration data sets while distinguishing between trip purposes.

We present both a flexible, easily adaptable method for calculating activity spaces of different shapes and a qualitative comparison of the four above-mentioned shape types on the basis of the given surveys. We can thus demonstrate that the choice of an appropriate shape representing an individual’s activity space is highly dependent on the spatial distributions and frequencies of the locations visited by the person in the given time period.

INTRODUCTION

Before Geographic Information Systems (GIS) were introduced, Euclidean measures like the standard deviational ellipse (SDE) and methods using place-based proxies for household locations (such as a zip-code centroid) were used to approximate the activity space of travelers, which is understood here as the area they have visited in person (see *(1)* for more details). However, most previous studies were restricted to samples of one- or two-day diaries because of the expense involved in data collection, especially in geocoding the addresses reported. With recent advances in GIS technology and through the increasing availability of spatially referenced data, activity space has become a more attractive tool for studying the spatial behavior of individuals. These same technological advances have enabled researchers to develop new measures of activity space that improve on the precision of the standard deviational ellipse and represent and analyze actual travel behavior more accurately.

The purpose of the micro-geographical concept of activity space is to capture the structures of the observed location choices of the individual traveler. A two-dimensional (geometric) form covers the places visited by an individual over a period of time in a defined manner. Previously, the standard deviation ellipse (SDE), the two-dimensional generalization of the confidence interval, was the preferred method because it easy to calculate (see *(1,2)* for further approaches). However, the SDE imposes a specific geometric form, which might or might not reflect the underlying behavior or urban form. In addition, it captures the underlying variance and therefore suggests rather large areas. Furthermore, if the SDE is calculated with respect to the home location, it imposes symmetry around the home even if one half of the area covered is free of locations visited. Alternatives avoiding these issues would therefore be desirable.

Schönfelder and Axhausen *(2)* suggested the shortest-path network or kernel-density-derived measures as alternatives. These are appealing, but they have computational drawbacks. In addition, shortest-path networks require substantial additional data in the form of a complete navigation network for the study area. Kernel-density-derived measures correlate very highly with the number of observed trips and add little new information. Nevertheless, they are a good semi-parametric way to map activity locations. If one accepts a particular geometry and a unique criterion to determine the parameters of this geometry a priori, then any geometric form could be used to capture the observed destinations: a circle, a square, a triangle, etc. In this paper four geometries will be tested, each reflecting a particular hypothesis about the form of a human activity space (see also *(3)* for a first implementation):

- the ellipse, which can capture an activity space with either one or two clusters of locations visited combined with a range of other locations outside these clusters. (One cluster can be captured because the ellipse can collapse to a circle);
- the Cassini oval, which also captures two clusters, but without intermediate locations between the clusters;
- the bean curve, which can accommodate three clusters, but is more flexible than a triangle;
- the superellipse, which comprises a circle and an ellipse, but can address a situation with four clusters by including a diamond-like form.

This paper presents a method for matching these geometries to observed activity locations. It was applied to five long-duration diaries or GPS observations. The method improved estimates of the size of human activity space and for the first time also provided insights into its structure. The paper focuses on the algorithmic implementation of the calculation of activity spaces of the above-defined shapes. Furthermore, it is our objective to separate the calculation of an activity space from its actual shape.

After a brief overview of the literature covering work from biology, geography, and planning, we shall present the algorithm implemented, separated into the optimization process and the activity shapes defined above. The final section describes the data sets and the estimation results. The paper concludes with suggestions for further research.

OVERVIEW OF THE LITERATURE

One of the first aggregate approaches to estimating an individual's range of movement and contact was Hägerstrand's Mean Information Field *(4)*. Since Hägerstrand and his colleagues could not use longitudinal movement information which would have fulfilled the research requirements, they used local migration data to test the model. The concept was later applied to other data sources and in different contexts, interestingly also to one of the first longitudinal travel data sets ever, the Cedar Rapids movement study data *(5, 6)*.

Lynch's work *(7)* was based on the assumption that the perception of space is a highly subjective process, which contrasts with the generalized representation of space in cartography. Motivated by Lynch's interest in the relationship between the structures as well as in the quality of architecture and human perception, Lynch found out that the mental maps of individuals, i.e., the image which human beings develop about their (travel) environment, are

- more or less biased;
- are simplifications of the real world;
- are group-specific; and
- are composed of about five basic elements which have different meanings for the structure of urban space in different cities (paths, border lines, areas, foci, and landmarks).

Mental maps mainly act as individualized cognitive support for spatial ordering and orientation. Mental maps and their formation may be captured only through indirect methods. Lynch used memory protocols and—as a main approach—map sketches by test persons.

Inspired by Hägerstrand's space-time paths *(8)*, Lenntorp *(9)* developed the concept of space-time prisms. He operationalized Hägerstrand's ideas about measuring individual accessibility based on the notion of a person's reach. Space-time prisms define the possible locations for a space-time path with obligatory activities such as work fixing the shape of the prism by predefining the person's location.

Finally, the activity-space concept, which was developed in the 1960s and 1970s in parallel with several of the approaches presented above to describe individual perception, knowledge, and actual usage of space (see *(10)* for a discussion), is used to represent the space which contains the places frequented by an individual over a period of time. Activity spaces are (geometric) indicators of observed or actual daily travel patterns (see also *(11)* for more details). This is stressed here because related concepts such as action space, perceptual space, mental maps or space-time prisms mainly describe an individual's travel potential.

Activity spaces are subject to fundamental geographical principles such as distance decay and directional bias, which implies that the probability of (regular) contact with a location usually decreases with its distance from the peg(s) of daily life (particularly home) and the necessary deviation from the main orientation or direction of daily travel. The latter refers to preferences for a particular place over other places of equal or similar distance due to some perceived quality of the preferred place *(12)*.

Newsome, Walcott and Smith *(13)* presented the use of ellipses to portray activity spaces and offered one example of how the ellipse construct can be used to analyze urban travel characteristics, based on observed trip-making behavior and socio-economic variables resulting in the importance of home location and household size for determining elliptic activity-space characteristics.

THE ACTIVITY-SPACE CALCULATION ALGORITHM

The algorithm presented in this paper is an element of the (MATSim-T) multi-agent transportation simulation toolkit *(14)*. The toolkit is based on a well-defined database describing a given scenario which consists of spatial data, transportation networks, survey information, and detailed descriptions of each individual active in the scenario (see Figure 1). Each optional element reads the defined XML (eXtensible Markup Language) data, stores it in an appropriate data structure, and writes it again in an enriched, reduced, or even unchanged form in the XML-data format. The person data structure is based on the MATSim-T daily schedule DTD (in MATSim-T called a plans-DTD) and is used as a working file to enrich a person's description, her or his personal knowledge of the scenario, and what she or he plans to do. In the minimum version the file holds only the identity number (ID) of all persons modeled. However, it is possible to add a large amount of information about each person to the file, such as age, sex, car ownership, home, work, other locations visited by the individual, etc. One part of the data structure actually defines *personal activity spaces*. These data points will be calculated by the activity-space calculation algorithm. Algorithms such as the one described in this paper can be added to each package to verify, manipulate, add, or delete data items according to the purpose of the algorithm. The flexible use of MATSim-T allows one to add other types of activity spaces than the ones implemented in here. Providing MATSim-T with the activity-space calculation algorithm allows one to use the result to analyze various behavioral characteristics of individuals for a given survey.

The simplex optimization technique

The Nelder-Mead simplex optimization technique *(15)* is an algorithm for finding the local minimum of a function of several variables. For two variables, a simplex is a triangle, starting with three two-dimensional points. These three points correspond to the three vertices of a triangle and constitute the first simplex. The method is a pattern search that compares function values at the three vertices of the triangle. The worst vertex, where $f(x,y)$ is largest, is rejected and replaced with a new vertex. A new triangle is formed and the search is continued (see Figure 2). The process generates a sequence of triangles for which the function values at the vertices get smaller and smaller. The size of the triangles is reduced and the coordinates of the minimum point are found. Similarly, for a three-dimensional space, four initial observations are required, thus defining a tetrahedral body. In general, for a function of n variables, the algorithm maintains a set of $n + 1$ points forming the vertices of a simplex in n-dimensional space. This simplex is successively updated at each iteration by discarding the vertex with the highest function value and replacing it with a new vertex that has a lower function value. Here is a brief outline of the steps involved in two-dimensional simplex optimization, following Mathews and Fink *(16)*.

Initial triangle BGW

Let the function to be minimized be $f(x,y)$. Since it is a two-variable function, we are dealing with two-dimensional parameter space and hence require three initial observations to serve as the three vertices of the triangle constituting the first simplex. Let these vertices be given by $V_k = (x_k,y_k)$, $k = 1,2,3$. The function $f(x,y)$ is then evaluated at each of the three points: $z_k = f(x_k,y_k)$ for $k = 1,2,3$. The points are then sorted according to the calculated response as the best, the next best, and the worst. We use the notation

$$B = (x_1, y_1),\ G = (x_2, y_2),\ W = (x_3, y_3),$$

where B is the best, G is the next to best, and W is the worst vertex (see Figure 2b).

Midpoint of the good side

The next step involves calculating the midpoint of the line segment joining B and G. It is found by averaging the coordinates as

$$M = \frac{(B+G)}{2} = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right).$$

Reflection using the test-point R

The function *f(x,y)* decreases as we move along the side of the triangle from *W* to either of the points *B* and *G*, implying that function takes on smaller values at points located away from *W* on the opposite side of the line between *B* and *G*. A test-point *R* is chosen by "reflecting" the triangle through edge *BG*. In order to determine *R*, the midpoint of edge *BG* is calculated first. Then a line segment is drawn from *W* to *M*. Let its length be *d*. This line segment is extended a distance *d* through *M* to locate the point *R*. The vector formula of *R* is

$$R = M + (M - W) = 2M - W.$$

Expansion using the point E

If the value of the function calculated at *R* is smaller than that at *W*, it signifies that we have moved in the correct direction towards the minimum. However, since this is not the minimal point, we can try to achieve a better approximation of it by extending the line segment through *M* and *R* to the point *E* (see Figure 2b), forming an extended triangle *BGE*. The point *E* is found in the same manner as *R* by moving an additional distance *d* along the line joining *M* and *R*. If the value of the function is lower than that at *R*, it implies that *E* is a better vertex than *R*. The vector formula of *E* is given by

$$E = R + (R - M) = 2R - M.$$

Contraction using the point C

If the initial reflection fails, i.e., if the calculated value of the function at *R* is higher than that at *W* or if *R* is not within the accepted limits of the parameters, then a contraction is needed. A point *C*, called the "contracted point", is found by calculating the midpoint of the line segment joining *W* and *M*. In this case, the point *C* replaces *W* in the simplex. However, if the function value at *C* is not lower than that at *W*, the points *G* and *W* must be shrunk towards *B*. The point *G* is replaced with *M*, and *W* is replaced with *S*, which is the midpoint of the line segment joining *B* with *W* (see Figure 2c).

Therefore, at the end of every iteration step, we generate a new simplex with a set of three points ($n + 1$ points for n-dimensional simplex optimization). The process is repeated until we approach the optimum value or until the improvement of the response becomes insignificant between successive iterations.

Using simplex optimization to calculate activity spaces

The simplex-optimization algorithm described above can be used to find activity spaces of a specified shape such that the captured area will be minimized. Furthermore, we want to predefine a certain percentage of the given geocoded locations which the resulting activity space should cover. It is therefore crucial to define which parameters must be preset by the user, which ones define the parameter space in which the simplex algorithm optimizes, and—last but not least—we need to define the objective function which has to be minimized.

The input parameters that must be predefined by the user are:

- the type of the activity space to be minimized (ellipse, superellipse, Cassini oval or bean curve);
- the type of locations chosen for calculating the activity space (here we have used "home", "work", "education", "shop", "leisure", and also the category "all", which defines the activity space to include all given locations);
- the minimum coverage (*cover* = [0.0,1.0]) to define the percentage of locations which must be covered by the resulting activity space;

- the step size for angle θ, which defines the rotation of the shape based on the Euclidean coordinate system. The smaller the step size, the more precisely the resulting activity space will fit the global minimum.

Note that the angle θ could, in principal, be part of the simplex parameter space. However, during our research, we have discovered that the search space including θ leads to problems in the optimization technique. In other words, the solution space is distorted towards the direction of θ, which greatly diminishes the performance of the simplex algorithm.

The objective function $f(x_1,x_2,\ldots,x_n)$ calculates the captured area of the given activity space defined by its parameters $x_1,x_2,\ldots,x_n$. For all four activity spaces discussed in this paper we have used the objective parameters *x0* and *y0* to define the center of the activity space in the given Euclidean coordinate system. The horizontal and vertical extensions (*a* and *b*) are also part of the search space. But since we have already set the coverage *cover*, we are able to reduce the search space by one dimension: instead of calculating $f(x0,x1,a,b,\ldots,x_n)$, we reduce the search space so that it includes only activity spaces within the predefined *cover* value. Therefore, we substitute *a* and *b* by the ratio *ratio* = *b*/*a* and let the objective function calculate *a* and *b* so that it fulfills *ratio* = *b*/*a* and covers the defined amount of locations (preset by the *cover* parameter). As a result, the search space of the simplex algorithm is defined by the coordinates *x0* and *y0*, the *ratio* and, depending on the given shape of the activity space, additional parameters.

The objective function $f(x0,x1,a,b,\ldots,x_n)$ is therefore replaced by $f(x0,x1,ratio,\ldots,x_n)$. In the objective function, the actual extends of *a* and *b* will be calculated via the bisection method. A detailed description of the search space of each evaluated shape is described in the next section.

Geometries implemented

As discussed above, the geometries chosen impose a parametric form on the activity space. Although these geometries are an abstract representation of travelers' activity spaces, they actually allow for the construction of a more comprehensive and realistic picture of travel behavior than has previously been provided. They provide an understandable graphical representation of a combination of elusive concepts, and they provide several appropriate measures of quantification and comparability to allow the further analysis of activity spaces. This section presents the mathematical definition of the geometries (see Figure 3 for their graphical representations).

The ellipse

An ellipse is defined as the locus of points *P*, such that the sum of the distances from *P* to two fixed points F_1 and F_2 (called foci) is constant. That is, distances PF_1 and PF_2 are equal to 2*a*, where *a* is a positive constant. An ellipse centered at the origin of an Euclidian coordinate system with its major axis along the abscissa is defined by the equation of the elliptical object:

$$\left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2=1 .$$

The same ellipse is also represented by the parametric equations

$$x=a\cos t;\quad y=b\sin t ,$$

where $t = [0,2\pi]$.

If the ellipse is not centered at the origin of the Euclidian coordinate system it may be defined as

$$\left(\frac{x-x0}{a}\right)^2+\left(\frac{y-y0}{b}\right)^2=1 ,$$

where (*x0,y0*) is the center.

The parametric form of an ellipse centered at (*x0,y0*) and rotated through an angle θ is given by

$$x = a\cos t\cos\theta - b\sin t\sin\theta + x0, \quad y = a\cos t\sin\theta + b\sin t\cos\theta + y0.$$

The area of an ellipse is given by

$$A = \pi \times a \times b.$$

In order to produce an ellipse covering 95% of the activity locations, the following four parameters are used:

- the x-coordinate of the center of the ellipse (*x0*);
- the y-coordinate of the center of the ellipse (*y0*);
- the orientation of the major axis of the ellipse (θ); and
- the ratio (*ratio*) of the length of the semi-minor axis (*b*) to the length of the semi-major axis (*a*) of the ellipse.

Superellipse

A superellipse (Lamé curve), centered at origin, is defined in the Cartesian coordinate system as the set of points satisfying the equation

$$\left|\frac{x}{a}\right|^r + \left|\frac{y}{b}\right|^r = 1,$$

where $r > 0$ and *a* and *b* are the radii of the oval shape. The case $r = 2$ yields an ordinary ellipse; *r* values less than 2 result in hyperellipses with pointed corners in the *x* and *y* directions resembling crosses; *r* values greater than 2 yield hyperellipses which increasingly resemble rectangles.

The superellipse may be described parametrically as

$$x = a\cos^{2/r} t, \quad y = b\sin^{2/r} t.$$

Its center *(x0,y0)* is given by

$$\left|\frac{x - x0}{a}\right|^r + \left|\frac{y - y0}{b}\right|^r = 1,$$

with the parametric equations

$$x = a\cos^{2/r} t + x0, \quad y = b\sin^{2/r} t + y0.$$

A superellipse with its center at *(x0,y0)* and an orientation of θ with the x-axis is defined parametrically as

$$x = a\cos^{2/r} t\cos\theta - b\sin^{2/r} t\sin\theta + x0, \quad y = a\cos^{2/r} t\sin\theta + b\sin^{2/r} t\cos\theta + y0.$$

The area of the superellipse is given by

$$A = \frac{4^{1-1/r} ab\sqrt{\pi} * \Gamma(1+1/r)}{\Gamma(1/2+1/r)},$$

where Γ is the gamma function.

To construct an optimal superellipse, the following parameters must be calculated:

- the x-coordinate of the center of the superellipse (*x0*);
- the y-coordinate of the center of the superellipse (*y0*);
- the orientation of the major axis of the superellipse (θ);
- the ratio (*ratio*) of the length of the semi-minor axis (*b*) to the length of the semi-major axis (*a*) of the superellipse; and
- the exponent *r*.

The exponent *r* determines whether a superellipse will look like a diamond ($0 < r < 1$),a rectangle ($r = 1$), or more like an ellipse ($r > 1$). In this paper we are interested in diamond shapes; therefore, we shall reduce the search space of the simplex algorithm to values of *r* between zero and one.

The Cassini oval

A Cassini oval is defined as a locus of a point such that the product of its distances from two fixed points at a distance of 2*a* apart is a constant b^2.It can be expressed as

$$\left((x-a)^2+y^2\right)\left((x+a)^2+y^2\right)=b^4 .$$

The parametric form of the above curve is

$$x = a\cos t\sqrt{\cos 2t+\sqrt{\left(\frac{b}{a}\right)^4-\left(\frac{\sin 2t}{a}\right)^2}}, \quad y = a\sin t\sqrt{\cos 2t+\sqrt{\left(\frac{b}{a}\right)^4-\left(\frac{\sin 2t}{a}\right)^2}} .$$

The parametric equation of a Cassini oval at an angle θ and centered at ($x0$,$y0$) is

$$x=a\cos(\theta+t)\sqrt{\cos 2t+\sqrt{\left(\frac{b}{a}\right)^4-\left(\frac{\sin 2t}{a}\right)^2}}+x0;\ y=a\sin(\theta+t)\sqrt{\cos 2t+\sqrt{\left(\frac{b}{a}\right)^4-\left(\frac{\sin 2t}{a}\right)^2}}+y0 .$$

The area of the curve is given by

$$A=\frac{1}{2}r^2d\theta=\int_{-\Pi/4}^{\Pi/4}a^2\left[\cos(2\theta)\pm\sqrt{\left(\frac{b}{a}\right)^4-\sin^2(2\theta)}\right]d\theta .$$

The shape of the oval depends on the ratio *b*/*a*. When *b*/*a* is greater than 1, the locus is a single, connected loop. When *b*/*a* is less than 1, the locus comprises two separate parts, as shown in Figure 3. When *b*/*a* is equal to 1, the locus is a lemniscates. Since the situation *b*/*a* is less than 1 is not relevant to our purpose, it was excluded by constraining *b*/*a* to be greater than 1 throughout the optimization process.

For the Cassini oval, the optimization process is determined by the following four parameters:

- the x-coordinate of the center of the Cassini oval (*x0*);
- the y-coordinate of the center of the Cassini oval (*y0*);
- the orientation of the major axis of the Cassini oval (θ); and
- the ratio (*ratio*) of the length of the semi-minor axis (*b*) to the length of the semi-major axis (*a*) of the Cassini oval.

The bean curve

The standard unit bean curve, situated in the first and fourth quadrants, with its origin at one end and the horizontal axis of unit length oriented along the abscissa, is given by the equation

$$x^4+x^2y^2+y^4=x\left(x^2+y^2\right).$$

The parametric form of the bean curve is

$$x=\sin^2\theta, \quad y=\sin\theta\sqrt{\frac{\cos^2\theta+\cos\theta\sqrt{1+3\sin^2\theta}}{2}} .$$

The area of the bean curve is given by

$$A=\sqrt{2}\int_0^1\sqrt{x\left(1-x+\sqrt{1+(2-3x)x}\right)}dx\cdot a\cdot b\simeq 1.058049*a*b .$$

The optimization process is performed in a four-parameter space consisting of the following:

- the x-coordinate of the center of the bean (*x0*);
- the y-coordinate of the center of the bean (*y0*);
- the orientation of the major axis of the bean (θ); and
- the ratio (*ratio*), where *a* and *b* are the multiples by which the standard unit bean curve is stretched in the horizontal and vertical directions.

RESULTS AND DISCUSSION

Data sources

In the past, transport planners collected short-duration, generally one-day diaries because of the cost and effort required for long-duration studies. However, with the availability of the GPS-based tracking system and increased interest in studying travel behavior, a number of long-duration surveys and observational studies have become available.

The data sets employed here are the Uppsala five-week diary (*17*), the Mobidrive six-week diary (*18*), the ISA Rättfart GPS observational study (*19*, *20*), the SVI Stabilität six-week diary (*21*), and the AKTA GPS observational study (*22*).

The Uppsala survey was conducted in the city of Uppsala, Sweden, located approximately 70 km northwest of Stockholm, in the year 1971. A random sample of 20% of the population was drawn. The final sample size was 278 households with 488 persons, of which 92 households were chosen for further analysis. The addresses of all trips were geocoded by hand.

The Mobidrive survey, conducted over six years in the German cities of Halle/Saale and Karlsruhe in 1999, involved a total of 317 persons in 139 households. The trip destination addresses of all main study trips were geocoded, whereby the geocoding was positive for more than 98% of the trips. Small street blocks were used as the basis for geocoding the street addresses in Karlsruhe and Halle. However, addresses outside the urban areas were geocoded only on the basis of the centroid of the municipality.

The ISA Rättfart GPS study, carried out in the town of Borlänge in central Sweden, consists of movement information for vehicles which was collected fully automatically for up to two years. The vehicles were equipped with an on-board data-collection system consisting of a GPS receiver, a data-storage device running a GIS to map all movements, and a mobile power supply. The study was conducted from 1999 to 2002, involving more than 200 private and commercial cars equipped for periods of up to two years each.

The SVI Stabilität survey was performed in the Swiss canton of Thurgau in 2003, covering a six-week reporting period involving 99 households and 230 persons. Nearly all destination addresses and household locations could be geocoded with great precision.

In the AKTA study, carried out in the greater Copenhagen area of Denmark, approximately 400 cars were equipped with a GPS-based device during two 8 to 10-week periods in 2001 and 2002. Vehicle-movement data was collected every second. A telephone-based before-and-after survey, which consisted of attitude questions, accompanied the GPS-monitoring process. The available sample involved 50 vehicles and persons with 44 to 135 reported days and 125 to 1044 reported trips each. A trip was defined as beginning when the first satellite signal was received and ending when the engine was switched off. The locations visited were identified by a simple clustering technique which grouped adjacent trip ends into clusters using the SAS Fastclus procedure (*23*).

Setup

For all the surveys given above we have calculated a 95% coverage of activity spaces (*cover* = 0.95). It should be noted that it is generally not possible to obtain exactly the given

coverage because the number of locations excluded are discrete numbers. Therefore, the algorithm has calculated the lowest possible coverage above 0.95.

The frequency of visits to each location in the above studies is known. We used this as the weight of the location simply by multiplying the location by the frequency of visits. Thus, locations which were visited often are typically part of the calculated activity space.

We assigned an orientation step size of θ by θ_{step} = 22.5°. This was a good trade-off between computation time and the accuracy of the resulting activity space.

Results

Before discussing the results for all of the samples, a set of example results will be discussed and mapped. Figure 4 depicts the different optimal geometries covering 95% of the shopping locations visited by a respondent drawn from the "SVI Stabilität" study. The person reported 218 trips, sorted by the following activity types:

- 44 leisure activities at 24 unique leisure locations;
- 77 work activities at 11 unique work locations;
- 84 home activities at 1 unique home location;
- 12 shopping activities at 10 unique shop locations; and
- 1 education activity at 1 unique education location.

The point to be noted here is that the superellipse produced is not the most optimal one possible. This is due to the limited number of angles which were tested during the optimization process (see above). A smaller step size for angles could be adopted for higher accuracy, but at the cost of increased computation time.

Figure 5 shows the optimal bean curves for all locations as well as for the four activity-purpose subsets. The bean curve covering 95% of all activity locations is smaller than that for leisure activities, because there are peripheral and rarely-visited leisure locations.

The optimal parameters were defined for all of the proposed geometries and all five data sets (Table 1). In addition, the best, i.e., smallest, of the four geometries was determined for each respondent. All distributions are highly skewed to the left with substantially smaller medians than means. The large coefficients of variation underline this variance, while the mean and median of the best geometry (meaning that we counted the best activity-space type for each single individual) is much smaller than any of the other distributions. Therefore, none of the given geometries is appropriate for all respondents. The Cassini oval fits in about two of three cases, but not always. This indicates that the Cassini oval best reflects the typical pattern of two dominant clusters of activity (*1*): mostly home, work, and education, but also frequently a leisure-oriented cluster. The bean curve captures a pattern with three clusters in less than 10% of the cases.

If we compare the number of occurrences of a specific ratio of one of the given activity spaces, as shown in Figure 6, there are several notable aspects:

1. In the ellipse, superellipse, and bean-curve histograms, there is a high amount of ratio (b/a) almost equal to zero. There are artifacts created by calculating activity spaces for one or two geocoded locations. Since those activity spaces do not have any use, we do not consider them in the analysis.
2. The ratio histogram of the ellipse shows a more or less normal distribution with the mean of ratio b/a equal to 1. In other words, many resulting ellipse activity spaces capture an area similar to a circle. With the information given in table 1, we already know that there are many exceptions in which the ellipse type does not cover an appropriate area while other shapes minimize the space better.
3. Similar conclusions can be drawn about the superellipse. While the ratio is normally distributed around b/a = 1, the exponent (r) is either very small (which forms a very

spiky-shaped diamond) or almost 1 (diamond-shaped). Since the results in Table 1 show that the superellipse does not capture an appropriate activity space compared to the other shapes, it is no improvement on the ellipse shape.

4. By contrast, the Cassini oval delivered very promising output. Most parts of the shapes are lemniscates. The others tend to form a circle. This indicates that the Cassini oval nicely captures location sets which are concentrated at two spatial regions. The ellipse and the superellipse fail in such cases. The Cassini oval—per definition—consumes a small amount of space compared to the ellipse. Therefore, it is not surprising that the Cassini oval has a high frequency of being the best solution.
5. Last but not least, there is also a promising histogram for the bean curve. Since its shape is the most flexible of the four ones tested, we could capture locations sets with two groups ($b/a < 1$) as well as with three groups ($b/a > 1$). Circle-like activity spaces could also be captured ($b/a \sim 1$). Two-group location sets are captured better by the Cassini oval, but the bean curve is better trimmed than the ellipse.

SUMMARY AND OUTLOOK

In this paper we have presented a generic activity-space calculation module embedded in the MATSim-T project. It has the flexibility to produce an appropriate activity-space representation for a given location set. The module can be varied according to the type of shape of the activity space, the type of location to be covered by the activity space, and the percentage of locations which the activity space should cover. The module can be used in MATSim-T to enrich the person description of the MATSim-DataBase. Therefore, it provides a clean interface for analyzing or answering spatial-behavior questions of almost any kind.

The results in the previous section show that none of the defined types of shape for an activity space is always the best activity space representation for all individuals. It highly depends on the spatial distribution and frequency of visits to the locations by the person. The Cassini oval typically best fits individuals with two clusters of locations (e.g., a daily commuter), while the bean curve would, for example, be a good activity space for describing weekly commuters (with a work location, a weekday home location, and a weekend home location cluster). The results also clearly show that the—often used—ellipse seldom produces a good representation of an activity space. Actually, it is only useful if the visited locations are spread out more or less randomly. That is why in such cases the ellipse shape is close to a circle.

It would be desirable to test other, more enhanced shapes (i.e., a combination of the Cassini oval and the bean curve), since all of the tested shapes had their limitations. We may want to implement shapes which are less specialized and are able to define activity space much more efficiently and accurately.

The underlying simplex algorithm fulfills its purpose only partially. It is dissatisfying to preset the orientation of the shape stepwise and then pick the best solution. Some tests that involved adding the orientation into the search space clearly showed that the simplex algorithm could get stuck at a local optimum. To gain more flexibility (and computational speed), it is necessary to replace the simplex algorithm by a more stable optimization model which can handle more complex search spaces. Evolutionary strategies like the covariance matrix adaptation (CMA) used in (*10*) are fairly good approaches.

It would be interesting to discover what factors influence the activity space sizes of different individuals. A detailed statistical analysis needs to be performed to see if socio-demographic variables affect individual travel behavior and thus individual activity space.

The empirical results reported in this paper suggest that the activity-space concept has the potential of becoming a more widely-used tool for studying spatial access, transport policy, and planning and can be used to understand urban travel behavior. It can be used to

evaluate present and future urban structure and accordingly come up with solutions to satisfy the activity demand in a household's neighborhood, resulting in reduced travel expenses, congestion, and emissions. With the growing recognition of the need for long-duration data sets, future analyses will lead to improved understanding of urban travel behavior and thus more accurate forecasts and enhanced policy analyses.

ACKNOWLEDGEMENTS

The Uppsala travel diary data was made available by Prof. Susan Hanson, Clark University, Worcester. The ISA Rättfart GPS study data set was made available by transport psychologists from the universities of Dalarna and Uppsala. The Centre for Traffic and Transport of the Technical University of Denmark kindly provided the Copenhagen AKTA data.

REFERENCES

1. Schönfelder, S. Urban Rhythms, PhD dissertation, ETH Zurich, Zurich, 2006.
2. Schönfelder, S. and K.W. Axhausen. Activity spaces: Measures of Social Exclusion? *Transport Policy, Vol. 10, No. 4*, 2005, pp.273-286.
3. Vaze V.S., S. Schönfelder and K.W. Axhausen. Optimization of Continuous Space Representation for Human Activity Spaces, *Tech. Rep., No. 295*, Institut für Verkehrsplanung and Transportsysteme (IVT), ETH Zurich, Zurich, 2005.
4. Hägerstrand, T. *Innovationsforloppet ur Korologisk Synpunkt*, Gleerup, Lund, 1965.
5. Garrison, W.L., B.J.L. Berry, D.F. Marble, J.D. Nystuen and R.L. Morrill. Studies of Highway Development and Geographic Change, University of Washington, Seattle, 1959.
6. Marble, D.F. and J.D. Nystuen. An Approach to the Direct Measurement of Community Mean Information Fields, *Papers and Proceedings of the Regional Science Association, No. 11***,** 1963, pp. 99-109.
7. Lynch, D. *The Image of the City*, MIT-Press, Cambridge, 1960.
8. Hägerstrand, T. What about People in Regional Science?, *Papers of the Regional Science Association, Vol. 24, No. 1*, 1970, pp. 7-21.
9. Lenntorp, B. Paths in Space-Time Environments: a Time-Geographic Study of the Movement Possibilities of Individuals*. Lund Studies in Geography, Series B, No. 44*, 1976.
10. Charypar, David, K.W. Axhausen and K. Nagel. Implementing Activity-Based Models: Accelerating the Replanning Process of Agents Using an Evolution Strategy, paper presented at the *11th International Conference on Travel Behaviour Research*, Kyoto, 2006.
11. Axhausen, K.W. A dynamic understanding of travel demand: A sketch, *Tech. Rep., No. 119*, IVT, ETH Zurich, Zurich, 2002.
12. Golledge, R.G. and R.J. Stimson, *Spatial Behavior*, Guilford Press, New York, 1997.
13. Newsome, T.H., W.A. Walcott and P.D. Smith. Urban Activity Spaces: Illustrations and Application of a Conceptual Model for Integrating the Time and Space Dimensions, *Transportation, Vol. 25, No. 4*, 1998, pp. 357-377.
14. Balmer, M., K.W. Axhausen and K. Nagel. A Demand-Generation Framework for Large-Scale Micro-Simulations*, TRB 85th Annual Meeting Compendium of Papers CD-ROM*, Transportation Research Board, Washington, D.C, 2006.
15. Nelder, J.A. and R. Mead. A simplex method for function minimization. *Computer Journal, No. 7***,** 1965, pp.308-313.
16. Mathews, J.H. and K.K. Fink. *Numerical Methods Using Matlab*, 4th Edition, Prentice Hall, Upper Saddle River, New Jersey, 2004.

17. Hanson, S. and K.O. Burnett. The Analysis of Travel as an Example of Complex Human Behaviour in Spatially-Constraint Situation: Definition and measurement issues, *Transportation Research A, Vol. 16, No. 2*, 1982, pp. 87-102.
18. Axhausen, K.W., A. Zimmermann, S. Schönfelder, G. Rindsfüser and T. Haupt. Observing the Rhythms of Daily Life: A Six-Week Travel Diary, *Transportation, Vol. 29, No. 2*, 2002, pp. 95-124.
19. Schönfelder, S., K.W. Axhausen, N. Antille and M. Bierlaire. Exploring the Potentials of Automatically Collected GPS Data for Travel Behaviour Analysis - A Swedish Data Source, in J. Möltgen and A. Wytzisk, eds. *GI-Technologien für Verkehr und Logistik, No. 13***,** 2002, pp. 155-179, Institut für Geoinformatik, University of Munster, Munster.
20. Schönfelder, S. and U. Samaga. Where do you want to go today? - More Observations on Daily Mobility, paper presented at the *3rd Swiss Transport Research Conference*, *Tech. Rep., No. 179*, Ascona, March 2003.
21. Axhausen K.W., M. Löchl, R. Schlich, T. Buhl and P. Widmer. Fatigue in Long-Duration Travel Diaries, *Tech. Rep., No. 286*, IVT, ETH Zurich, Zurich, 2004.
22. Nielsen, O.A. and G. Jovicic. The AKTA Road Pricing Experiment in Copenhagen, paper presented at the *10th International Conference on Travel Behaviour Research*, Lucerne, August 2003.
23. Anderberg, M.R. Cluster Analysis for Applications, *Academic Press*, New York, 1973.

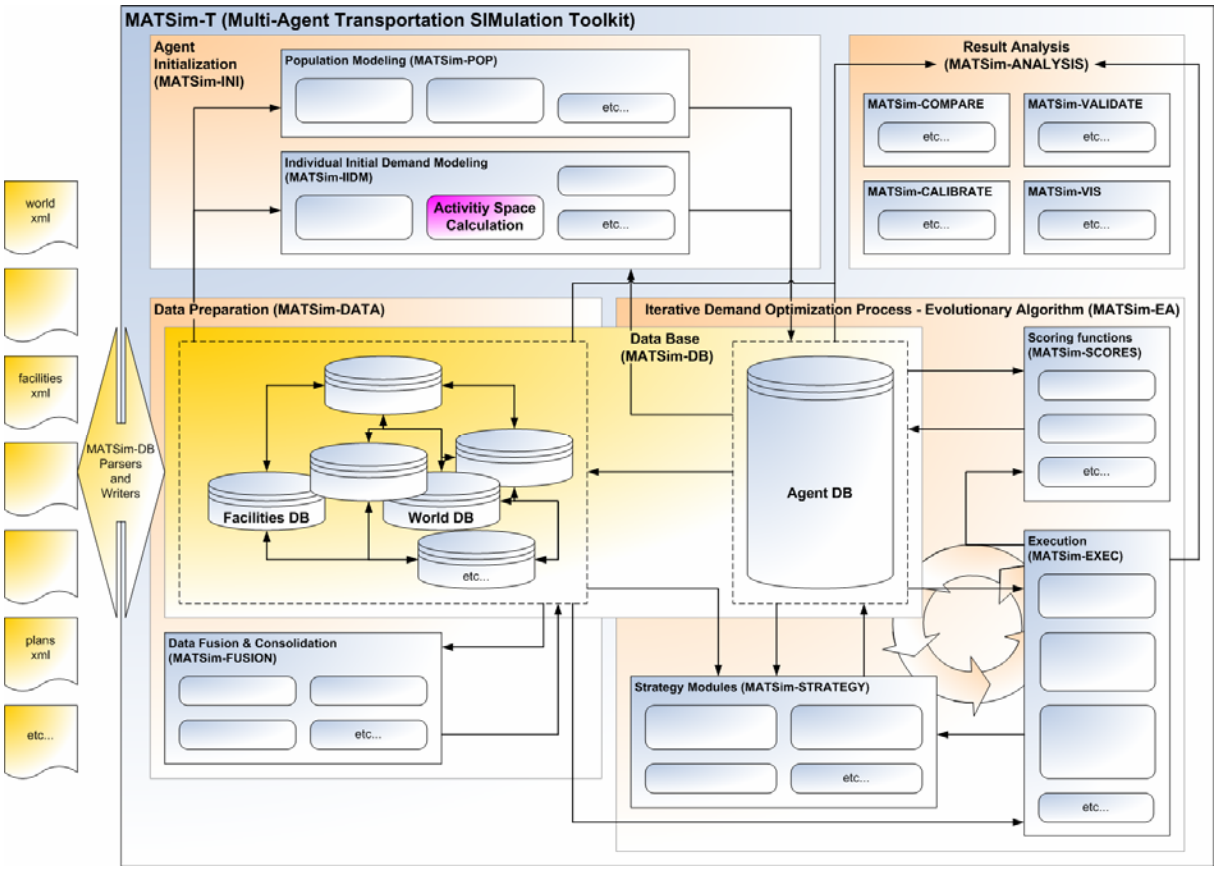

FIGURE 1: Schematic overview of MATSim-T

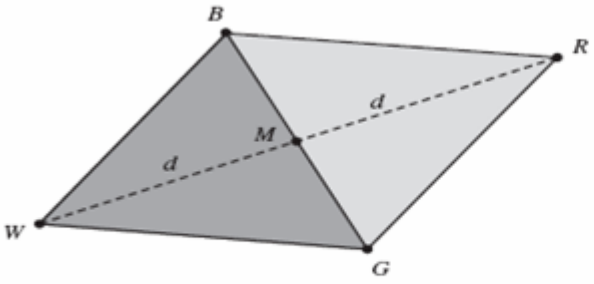

Figure 2a: The triangle *BGW*, midpoint *M*, and the reflected point *R*

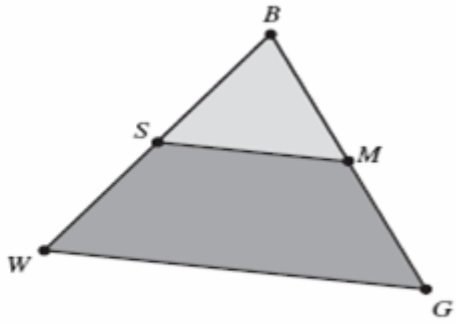

Figure 2c: Shrinking the triangle towards *B*

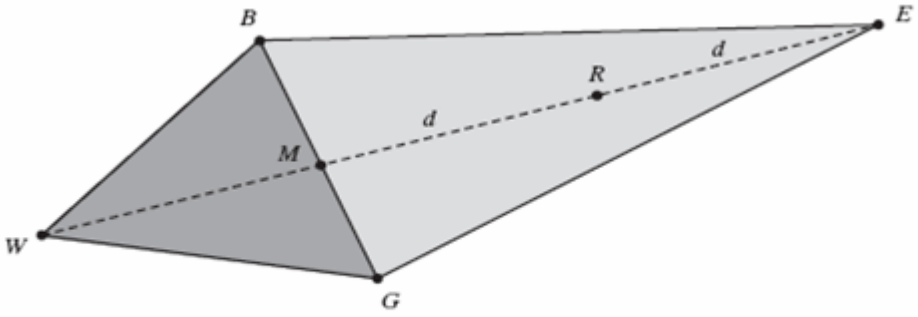

Figure 2b: The triangle *BGW*, point *R*, and the extended point *E*

FIGURE 2: Steps of the Nelder-Mead simplex algorithm

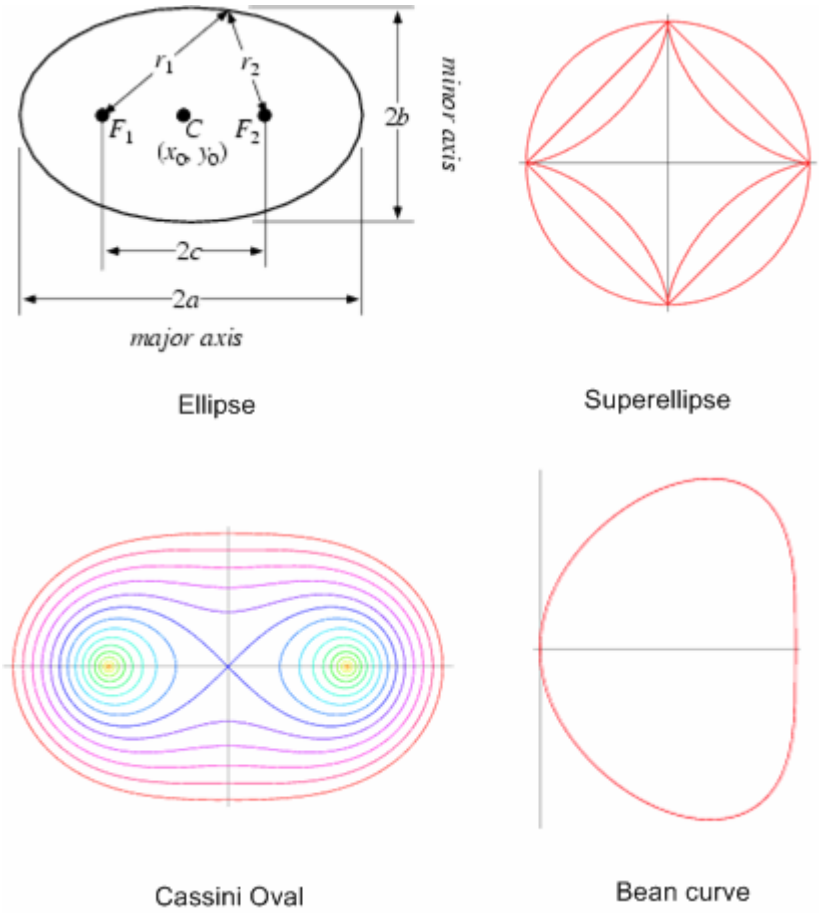

FIGURE 3: Geometries tested

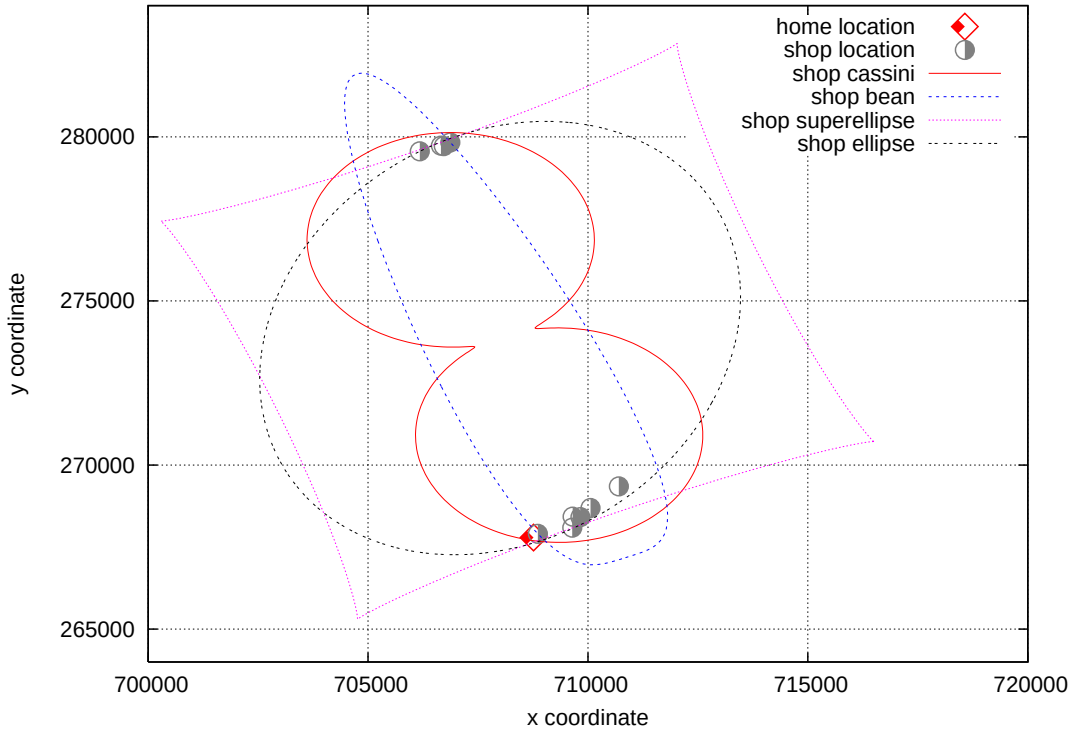

FIGURE 4: Example of optimal geometries for the shopping (plus home) locations

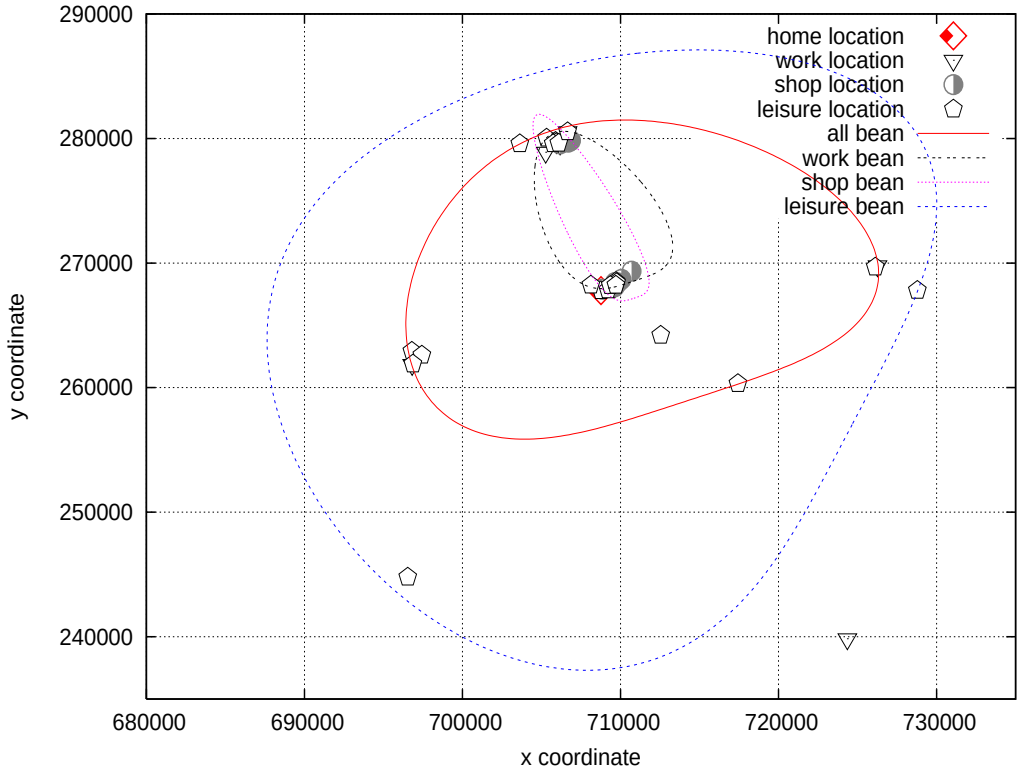

FIGURE 5: Example of 95% coverage of activity spaces by weighted purpose (bean curves)

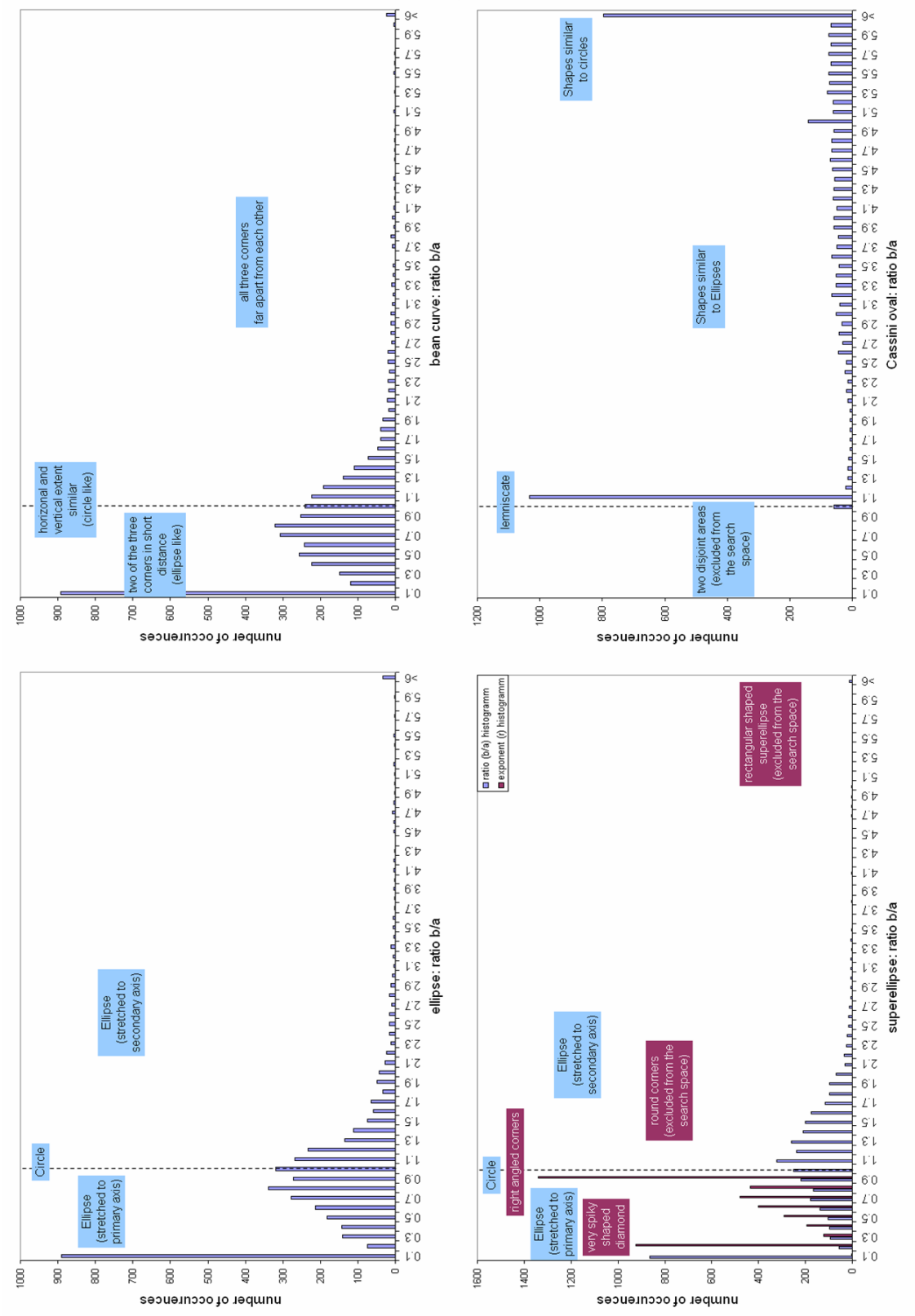

FIGURE 6: Occurrence distributions of ratio (b/a) for the four shapes

Uppsala diary	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	115.71	197.80	145.40	200.55	92.93
Median	9.04	16.10	13.70	18.40	8.96
Standard deviation	354.80	652.46	508.67	674.88	326.45
Frequency of being the best solution	113	7	10	14	144

Mobidrive diary, Halle / Karlsruhe	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	384.74	629.12	597.88	558.71	333.63
Median	42.30	71.50	66.70	76.90	37.60
Standard deviation	1909.28	3175.02	3166.75	2846.46	1665.40
Frequency of being the best solution	257	16	20	24	317

ISA Rättfart GPS study, Borlänge	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	102.10	141.20	165.43	147.98	88.79
Median	65.60	110.00	102.00	112.00	65.60
Standard deviation	128.44	110.23	214.70	126.55	92.48
Frequency of being the best solution	53	1	5	7	66

SVI Stabilität survey, Thurgau	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	1845.99	2819.77	2509.13	3419.68	1724.84
Median	347	598	460	637	319
Standard deviation	3893.65	6123.99	8087.50	3667.79	3219.41
Frequency of being the best solution	177	14	26	13	230

AKTA study, Copenhagen	Cassini	Ellipse	Bean curve	Superellipse	Best geometry
Mean	3322.98	4111.08	5013.48	4869.48	2603.34
Median	246.00	426.00	366.00	491.00	241.00
Standard deviation	7520.15	10729.63	11684.94	13688.16	6314.43
Frequency of being the best solution	146	26	17	12	201

TABLE 1: Optimal geometries for all locations: Summary of the results